Weekly Report 01/05/2014

Summary

For the new iListenPortal project, an initial prototype is deployed online. For the data cleaning project, I've figured out dynamic modeling techniques, and integrated Pandas as data manipulation package as well.
The soundscape project will start next Friday. We'll discuss over visualization proposals.

The VASA project

No VASA work.

The IListenPortal project

An initial prototype has been deployed (http://voxel.ecn.purdue.edu/soundscape/index.php). The project visualizes sound being uploaded via an app. Current version deals merely with complementary information of the sound (locations, contents and emotions related to the sound).

The data cleaning project

I sketched out a project schedule before paper submission (https://docs.google.com/spreadsheet/ccc?key=0AippZUxJvlStdE02UnlvOU1ZdldiNDl0bDFwSnpidmc&usp=drive_web#gid=0). My part is the server-end and Hu's part is the browser-end.
I also updated the implementation document, which extends some new thoughts and elaborates some implementation details meanwhile.
(https://docs.google.com/document/d/1aL9WIsHiTRqU60PO-i3PySfQCIIMEX-GWnsEYDUDS6w/edit)
New ideas are summarized below:

1. Users can edit raw data through both data table and visualization. The original editing process was designed as: data table selection → visualization (of selected columns) → data table edition (with finding revealed from visualization) → server-end data (update). But the two data table procedures sometimes can be redundant. Why not edit data directly visualization. For example, the value distribution of a certain attribute is visualized and users realize that some data are meaningless so they want to delete them. Instead of deleting from data table, it'll be more intuitive if users can delete them immediately from visualization and the server-end data updates correspondingly. I have to say that this is not just one step less, but data cleaning via visualization. And it's different from interactive data cleaning.
2. A second benefit of data cleaning via visualization is that visualization reveals correlations of attributes. Here the point is that users have to know attribute correlations before they take steps. They have to know as much of the data as possible, which is conflicted with the normal condition that usually they hardly know anything when they do data cleaning. And data cleaning is supposed to be a step towards better understanding of the data. Taking the same example of deleting meaningless data, the data might be meaningless in terms of one

attribute, but trivial in another. That is, the deletion might change data distribution in another attribute. Interactive data cleaning does not concern this issue, but visualization (I was thinking of multiple linked views or PCP) does. Or according to discussion with Ma, is it better to quantitatively evaluate the influence of every cleaning step in the workflow part maybe? It seems to be a task-specific data uncertainty work.

The elaboration includes overall GUI design and user interaction designs. So far, schema mapping still requires more discussions.

Future Work

1. Paper on visualization course teaching.
2. Data cleaning project: more design and more implementation.

P.S. Travel in Florida --- the state of sunshine

The places we went are Universal Studio, The Everglades National Park, and Key West. All highly recommended.


90 Miles to
CUBA
POINT
CONTINENTAL
U.S.A.


DO THE SPELL UNTIL